External quality assessment of multi-analyte chromatographic methods in oral fluid

Kristof Pil, Francis M Esposito, PhD, DABFT [b,] Alain Verstraete, MD, PhD [a,c]

In: Clinica Chimica Acta 2010; 411: 1041- 1045.

External quality assessment of multi-analyte chromatographic methods in oral fluid

Kristof Pil, MSc [a]

Francis M Esposito, PhD, DABFT [b]

Alain Verstraete, MD, PhD [a,c]

[a] *Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, De Pintelaan 185, Ghent, Belgium*

[b] *Center for Forensic Sciences, RTI International, 3040 Cornwallis Road, Research Triangle Park, North Carolina, 27709, USA*

[c] *Laboratory of Clinical Biology, Ghent University Hospital, De Pintelaan 185, Ghent, Belgium*

Grant nr: TREN-05-FP6TR-S07.61320-518404-DRUID (European Commission, 6th framework program)

Corresponding author:

Alain Verstraete

Ghent University

Department of Clinical Chemistry, Microbiology and Immunology

Policlinic 8, 2nd floor

De Pintelaan 185

B-9000 Ghent

Belgium

Tel +32 9 332 34 07

Fax +32 9 332 49 85

e-mail alain.verstraete@ugent.be

Abstract

Background: A proficiency testing scheme was set up for the DRUID (Driving under the influence of Drugs, Alcohol and Medicines) research project, funded by the European Commission, in which oral fluid is analysed by eleven laboratories. A common collection and analysis methodology is used: Statsure Saliva Sampler is used for collection and LC-MS/MS or GC-MS confirmation analysis of 22 substances is performed on all samples. Despite internal validation and quality control samples, external quality assessment is still necessary to further increase comparability of results. Four rounds of proficiency testing (PT) were organized between March 2008 and September 2009.

Methods: Qualitative results were evaluated using sensitivity and specificity. Quantitative results were evaluated using z-scores and the standard deviation of Horwitz.

Results: Specificity was above 99% in each round, sensitivity per analyte varied between 81.7 and 100%, 20 out of 22 analytes had a sensitivity above 90%. The percentage of satisfactory z-scores increased from 79.4% to 89.2%. This trend was seen for all drug classes, except zopiclone. Results were discussed with participating laboratories and problems were addressed.

Conclusions: Because of these corrective actions, DRUID laboratories have a lower variation in results than previously published PT schemes in oral fluid.

Keywords: oral fluid, proficiency testing, drugs of abuse, benzodiazepines, quality control

1. Introduction

The use of oral fluid as a matrix for detection of drugs of abuse has increased continuously over the last ten years.[1] More and more information has become available on the pharmacokinetics of both licit and illicit drugs[2], leading to a better understanding of the relationship between concentrations found in oral fluid and the effects of the drugs. Because of its non-invasive, supervised and easy collection, policy makers are also interested in oral fluid as an alternative matrix to blood and urine for detection of impaired drivers. Several countries worldwide (Australia, France, Belgium) have already implemented legislation based on oral fluid.

DRUID (Driving Under the Influence of Drugs, Alcohol and Medicines) is an integrated research project funded by the European Commission that deals with licit and illicit drugs in driving. Oral fluid is collected in this study for several purposes: it is taken from randomly selected drivers in order to calculate the prevalence of psychoactive substances in the general European driving population and to calculate odds ratios to be injured in an accident after taking these substances. Also, all experimental studies in DRUID are collecting oral fluid for comparison with concentrations in other matrices. Finally, oral fluid is taken as reference samples from patients in pain clinics, coffeeshops and centers for treatment of drug addiction to test rapid on-site testing devices for screening of drugs of abuse in oral fluid.

In total over 40,000 samples will be collected using the Statsure saliva sampler™ (Statsure Diagnostic Systems, Framingham, MA, USA). This collection device was chosen from ten devices based on ease of collection, stability and recovery of analytes and collection volume.[3] All samples have to be analyzed for the presence of a ‘core list’ of 22 substances containing both licit and illicit drugs. Minimum analytical cut-offs were defined for all analytes.[4] (Table 1) Samples are analyzed by eleven countries across Europe. Since results of the different laboratories will be combined for calculations and most labs had to develop new methods for this project, interlaboratory quality control (i.e. proficiency testing) is necessary. The aim of the proficiency-testing scheme was to ensure that the quality and comparability of the results from all laboratories. In the past, similar programs have been organized: UKNEQAS (United Kingdom National External Quality Assessment Scheme) was the first to publish interlaboratory results[5], followed by ORALVEQ[6, 7] and the ROSITA-2 project.[8] None of

these existing programs could however be used for the DRUID project since tests had to be adapted to the DRUID core list of analytes and cut-offs.

2. Materials and methods

Drug standards used in the preparation of oral fluid PT samples were purchased as powders from Sigma-Aldrich (St. Louis, Missouri, USA) or as methanolic solutions from Cerilliant (Round Rock, Texas, USA).

Oral fluid PT samples were prepared in a synthetic oral fluid matrix developed at RTI International. The synthetic oral fluid consisted of salts and proteins found in human oral fluid with no preservatives added. Each sample was formulated to contain 3 to 5 analytes. The 22 analytes, target concentration ranges and number of sample spikes for each analyte are listed in Table 1. 1.5 mL of neat oral fluid was dispensed into a 4 mL silanized amber vial (Supelco St. Louis, Missouri, USA), capped with a Teflon-lined cap (Supelco) and frozen until shipment. Samples for each survey year were prepared in a single production. Ten sample types were prepared in December 2007 for the two 2008 surveys (round 1 and 2) and ten new sample types were prepared in January 2009 for the two 2009 surveys round 3 and 4). Each sample vial was labeled with a unique identification number.

PT samples were packaged in cardboard boxes with styrofoam containers and dry ice (for countries able to accept dry ice and ice packs for those countries barring dry ice) in plastic buckets and shipped using express delivery to each of the participating laboratories. A total of 4 sets with 5 samples in each set were shipped between March 2008 and September 2009.

Laboratories were instructed to add 1 mL of the neat oral fluid sample to a StatSure collection device. The 1 mL of neat PT oral fluid was added directly to the buffer in the collection tube. The neat oral fluid was not to be added to the swab pad nor was the swab pad to be added to the buffer in the collection tube. Analytes were screened, identified and quantified using a mass spectroscopy-based technique. Reported analyte concentrations were corrected for dilutions to provide the concentration for the neat oral fluid shipped to the laboratory. Samples were expected to be tested and electronically reported to RTI within 10 working days after receipt. Results were reported back to each participating lab anonymously, but with identification to the DRUID coordinator to allow the latter to make corrective actions.

Qualitative results were evaluated using sensitivity and specificity. Sensitivity is defined as the number of analytes correctly reported positive divided by the total number of analytes spiked in the samples. Specificity is defined as the number of analytes correctly reported negative divided by the total number of core list analytes not spiked in the samples.

Quantification was evaluated using the standard deviation according to Horwitz (SD_{HOR}). This parameter is used in the IUPAC international Harmonized Protocol for proficiency testing and is independent of the results of the laboratories as a group. Z-scores were calculated using SD_{HOR}:

$$VC = 2^{(1-0,5\log C)}$$

VC = variation coefficient (%)

C = analyte concentration (kg/L)

$$z-score = \frac{result - target\ value}{SD_{HOR}}$$

3. Results

Eight laboratories reported results in the first round, three laboratories did not yet report because method development and validation were still being performed. In the second to fourth round, all eleven laboratories reported results. Not all laboratories reported results for ethanol, since in some countries ethanol concentration for each volunteer was already known based on breathalyzer results from standard police procedure and hence analysis was not mandatory.

Nine laboratories performed the analysis using LC-MS/MS (liquid chromatography tandem mass spectrometry), two laboratories used GC-MS (gas chromatography mass spectrometry). Ethanol analyses were performed with either headspace gas chromatography or enzymatic methods.

Average coefficients of variation (CV) were below 20% for 4 out of 22 analytes, 13 others had an average CV between 20 and 30%. Specificity was above 99% for all analytes; sensitivity varied between 81.7 and 100%, 20 out of 22 analytes had sensitivity above 90%. (Table 1).

In Figure 1, it can be seen that quantification improved over the rounds as well: laboratories with a satisfactory z-score (absolute value lower than 2) increased from 79.4% over 86.8% and 88.7% to 89.2%. Z-scores were also combined for the following groups: opiates (including methadone), benzodiazepines, amphetamines, cocaine (and benzoylecgonine), ethanol, delta-9-tetrahydrocannabinol (THC) and z-drugs (zolpidem and zopiclone). False negatives were mostly attributable to z-drugs, benzodiazepines and THC. Laboratories scored the lowest for quantification of amphetamines and z-drugs. For cocaine and opiates, almost no false negatives were observed while satisfactory quantification was achieved in respectively 89.4% and 92.5% of cases. Quantification of THC was satisfactory in 89.2% of cases. For ethanol, there were no false negatives results and satisfactory quantification was obtained in 95.8% of cases.

Results per substance were subdivided per testing round. These demonstrate that specificity for amphetamines, cocaine, opiates and THC improved over time, with no false negatives (except one FN for 6-acetylmorphine) remaining in round 3 and 4. (Figure 2 and Figure 3) For benzodiazepines as well, specificity improved over time, except in round 3 where one lab experienced temporary hardware problems. The Z-drugs remained the most difficult group during all rounds, although some improvement can be observed in round 4. (Figure 4) The majority of problems with the z-drugs were encountered with zopiclone: 12 out of 14 false negatives are for zopiclone, only 2 for zolpidem.

Overall, scores for quantification remained approximately the same over time, except for z-drugs and cocaine where a clear improvement was seen in round 4.

In testing rounds 1 and 4, several labs reported benzoylecgonine present in samples where only cocaine was spiked but no benzoylecgonine. After consultation with participating laboratories this problem could be explained by hydrolysis of cocaine during extraction and low extraction yields for benzoylecgonine or by pre-analytical hydrolysis, either in sample preparation at the manufacture or during shipment. Since the exact source of the benzoylecgonine could not be pointed out, and the presence of variable amounts of benzoylecgonine is to be expected in samples containing cocaine, these reported values were not scored as false positives.

4. Discussion

Prior to the DRUID project, most participating laboratories did not analyze oral fluid on a routine basis and therefore had to develop LC-MS/MS or GC-MS methods specifically for this project. In the first round of proficiency testing, most laboratories were therefore still in the process of development and validation or had only recently completed this, explaining the lower scores in the first rounds. After each round, results were discussed in meetings with the participating laboratories and the DRUID coordinator (based on Z-scores), problems were identified and solutions sought for. Thanks to this extensive distribution of knowledge between the participating laboratories, the quality of analyses increased significantly over time, leading to a last round with low false negative (0.9%) and false positive (0.4%) results. Satisfactory quantification was achieved in almost 90% of cases.

Zopiclone remained the most difficult compound for qualitative analysis. It is known that cyclopyrrolones can hydrolyse under different conditions to 2-amino-5-chloropyridine, which could explain the analytical problems.[9, 10] Moreover, when the participating laboratories developed their confirmation methods, deuterated zopiclone was not yet available to be used as internal standard. Amphetamines proved to be the most difficult class of drugs for quantitative analyses.

Laboratories were not allowed to analyze specimens until approval from the coordinator was obtained, which was based on internal validation and proficiency testing results. This was decided in order to further improve quality of analysis of specimens from DRUID volunteers.

Analysis using coefficients of variation shows that the DRUID laboratories have a lower variance in results than the participants of previously published proficiency testing programs for the majority of analytes: in this program the average CV was lower than 30% for 77.3% of analytes, while this was the case for only 46.1 % of analytes in UKNEQAS and none of the analytes in ORALVEQ.[5, 6]

It should be noted that two key variables in oral fluid analysis are not considered in the program: sample volume and recovery from the device. However for analysis of all real samples, variation is minimized since weighing of the sample volume is mandatory and recovery from the device is reproducible because of the use of the same collection and extraction device.

6. Acknowledgments

This work was carried out within the European Commission 6th framework programme funded integrated project DRUID. (contract no. TREN-05-FP6TR-S07.61320-518404-DRUID)

7. List of abbreviations

CV: coefficient of variation

DRUID: Driving Under the Influence of Drugs, alcohol and medicines

GC: Gas Chromatography

LC: Liquid Chromatography

MS: Mass Spectrometry

PT: proficiency testing

SD_{HOR}: standard deviation according to Horwitz

THC: delta-9-tetrahydrocannabinol

UKNEQAS: United Kingdom National External Quality Assessment Scheme

ROSITA: Roadside Testing Assessment

Table 1: number of samples spiked with target analytes, concentration ranges and DRUID cut-offs. Sensitivity and specificity, minimum, maximum and average coefficient of variation.

Analyte	DRUID cutoff (ng/mL)	Sample concentration range (ng/mL)	Number of Sample Challenges	Sens	Spec	CV			
						Average	Min	Max	Outliers[$]
6-acetylmorphine	5	8-15	4	93.9%	100%	27.4%	21.7%	34.8%	
Alprazolam	1	2-6	4	92.7%	100%	26.0%	17.4%	41.2%	2
Amphetamine	25	50-100	4	92.7%	100%	22.9%	12.3%	42.3%	
Benzoylecgonine	10	20-80	5	98.1%	100%	31.2%	25.5%	36.4%	2
Clonazepam	1	2-2	2	100%	100%	27.4%	21.3%	33.5%	
Cocaine	10	20-80	6	96.7%	99.3%	19.1%	13.1%	24.5%	
Codeine	20	25-60	3	100%	100%	35.3%	32.1%	41.1%	
Diazepam	5	10-15	2	100%	99.5%	22.2%	17.2%	27.2%	
Ethanol	0.1 g/L	0.2-0.8 g/L	8	100%	100%	11.2%	6.5%	18.9%	2
Flunitrazepam	1	2-6	2	81.8%	99.5%	28.8%	19.7%	37.9%	
Lorazepam	1	2-3	2	90.9%	100%	20.4%	18.1%	22.7%	1
MDMA	25	50-75	2	100%	100%	35.9%	31.2%	40.7%	
MDA	25	75-75	2	100%	100%	30.8%	26.9%	34.7%	
MDEA	25	75-75	2	100%	100%	28.9%	27.8%	30.0%	1
Methadone	20	30-40	2	100%	99.5%	18.5%	18.3%	18.7%	

Methamphetamine	25	40-75	3	94.7%	100%	28.7%	28.1%	29.0%	1
Morphine	20	25-60	4	100%	100%	28.1%	13.5%	46.6%	1
Nordiazepam	1	2-5	2	95.5%	99.5%	18.3%	16.2%	20.4%	
Oxazepam	5	10-10	2	94.7%	99.5%	27.4%	17.0%	37.7%	
THC	1	3-10	8	93.9%	100%	25.7%	13.7%	35.2%	1
Zolpidem	10	15-40	3	93.3%	100%	22.2%	16.2%	32.6%	
Zopiclone	10	15-40	6	81.7%	99.3%	34.5%	15.2%	65.1%	2

$ outlier defined as value with more than 100% deviation from mean

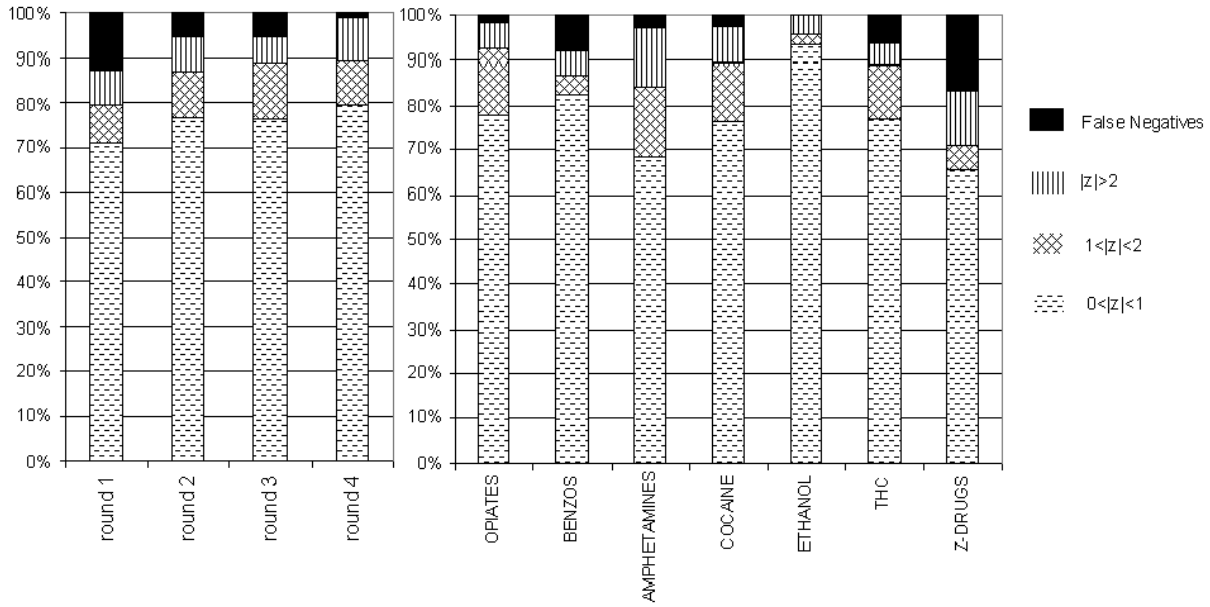

Figure 1: z-scores and false negatives for each testing round and per class of drugs (rounds combined)

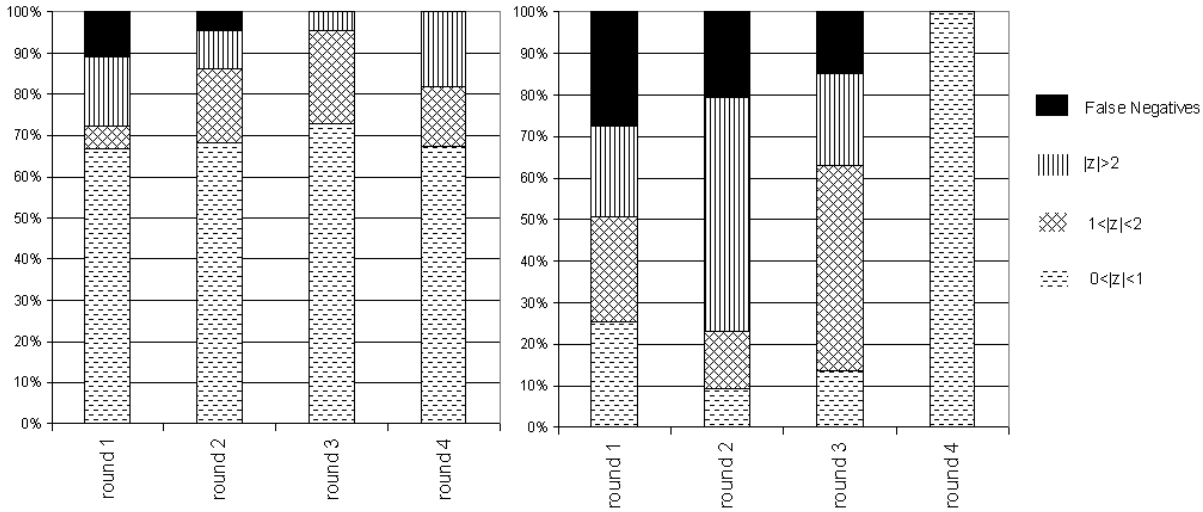

Figure 2: z-scores and false negatives for amphetamines (left) and cocaine (+benzoylecgonine) (right)

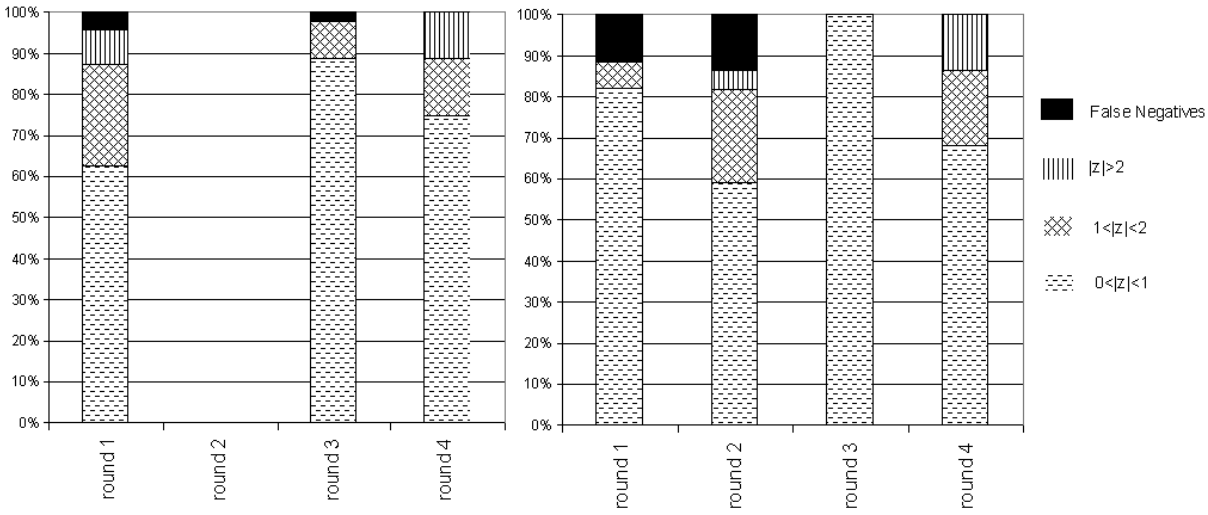

Figure 3: z-scores and false negatives for opiates (left; no opiates present in round 2) and THC (right)

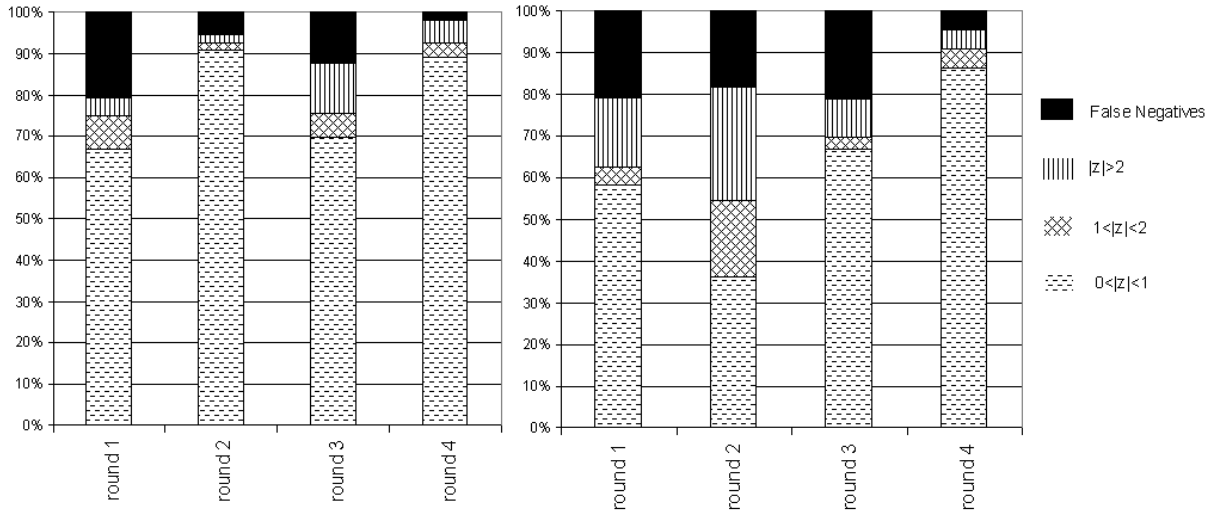

Figure 4: z-scores and false negatives for benzodiazepines (left), zopiclone and zolpidem (right)

Reference List

(1) Pil K, Verstraete A. Current Developments in Drug Testing in Oral Fluid. *Ther Drug Monit* 2008 April;30(2):196-202.

(2) Drummer OH. Review: Pharmacokinetics of illicit drugs in oral fluid. *Forensic Sci Int* 2005 June 10;150(2-3):133-42.

(3) Langel K, Engblom C, Pehrsson A, Gunnar T, Ariniemi K, Lillsunde P. Drug testing in oral fluid-evaluation of sample collection devices. *J Anal Toxicol* 2008 July;32(6):393-401.

(4) Pil K, Raes E, Verstraete A. The toxicological challenges in the European research project DRUID. *Forensic Science International Supplement Series* 2009; 1: 29-32.

(5) Clarke J, Wilson JF. Proficiency testing (external quality assessment) of drug detection in oral fluid. *Forensic Science International* 2005 June 10;150(2-3):161-4.

(6) Ventura M, Ventura R, Pichini S et al. ORALVEQ: external quality assessment scheme of drugs of abuse in oral fluid: results obtained in the first round performed in 2007. *Forensic Sci Int* 2008 November 20;182(1-3):35-40.

(7) Ventura M, Pichini S, Ventura R et al. Stability of drugs of abuse in oral fluid collection devices with purpose of external quality assessment schemes. *Ther Drug Monit* 2009 April;31(2):277-80.

(8) Wille SM, Raes E, Lillsunde P et al. Relationship between oral fluid and blood concentrations of drugs of abuse in drivers suspected of driving under the influence of drugs. *Ther Drug Monit* 2009 August;31(4):511-9.

(9) Fernandez C, Gimenez F, Mayrargue J, Thuillier A, Farinotti R. Degradation and racemization of zopiclone enantiomers in plasma and partially aqueous solutions. *Chirality* 1995;7(4):267-71.

(10) Mannaert E, Tytgat J, Daenens P. Detection of 2-amino-5-chloropyridine in urine as a parameter of zopiclone (Imovane) intake using HPLC with diode array detection. *J Anal Toxicol* 1997 May;21(3):208-12.